



A Corpus-based Analysis of Academic Near Synonyms : Consist, Comprise and Compose

Thitaporn Putklang*

Panupan Panthong**

Naviya Chutopama***

Received : May 12, 2024

Revised : August 2, 2024

Accepted : August 16, 2024

ABSTRACT

Researching academic vocabulary is crucial for understanding and effectively communicating within the academic community. However, second language learners may find synonyms confusing and prone to misuse. Therefore, this study aims to explore the differences in term of the degree of formality among three academic near synonyms—'consist,' 'comprise,' and 'compose'—and to investigate their noun and adverb collocations in relation to semantic preferences. The definitions of each target word were sourced from online dictionaries, while data on the frequency of each genre, collocation, and Mutual Information (MI) scores (≥ 3) were obtained from the COCA. The data analysis utilized Phoocharoensil's (2020a) framework. The results revealed that all three words are commonly used in the academic genre, even though 'compose' is not included in the AWL. Based on the analysis of semantic preferences, these three words cannot be substituted for one another. This study also suggests the pedagogical implication for the classroom, and teaching material construction.

Keywords : Corpus-based Study / Academic Vocabulary / Near Synonym / Coca / Semantic Preference

*Lecturer, English Program, Buriram Rajabhat University

**Lecturer, Business English Program, Buriram Rajabhat University

***Lecturer, English Program, Buriram Rajabhat University(Corresponding Author) e-mail: naviya.jc@bru.ac.th

Statement of Problems

The advent of corpus linguistics in this digital age has enabled numerous research studies to leverage these linguistic databases to contribute new knowledge and dynamically evolve the field of English Language Teaching (ELT). It is widely acknowledged that synonyms are one of the most pompous areas of vocabulary in the field of applied linguistics and ELT. Phoocharoensil (2022) identified two types of synonyms: absolute and near. The first type refers to the words mutually sharing the exact same meaning. Nonetheless, the number of such words in English is relatively rare (Taylor, 2003). Another type refers to a near synonym. The kind of word partially shares the similar meanings, yet it could be used interchangeably (Chaengchenkit, 2023; Chung, 2011; Narkprom, 2024; Panthong, et al., 2023). For example, 'advise', 'recommend', and 'suggest' – these three words convey the same meaning, but the collocations and grammatical patterns vary from word to word (Phitayakorn, 2016).

Such slight differences among words have sparked significant concerns among vocabulary experts. This is particularly important because second language learners often encounter challenges when learning and using near synonyms. These difficulties arise because such words, yet similar, cannot always be substituted for one another without altering the meaning or appropriateness of a sentence (Kruawong & Phoocharoensil, 2022; Lertcharoenwanich, 2023; Lertcharoenwanich & Phoocharoensil, 2022; Niwesworakarn, et al., 2023; Petcharat & Phoocharoensil, 2020; Phoocharoensil, 2020a, 2020b, 2021a, 2021b, 2022; Phoocharoensil & Kanokpermpoon, 2021; Sridhanyarat & Phoocharoensil, 2023). Researching near synonyms is therefore crucial, as it can shed light on how these words vary across different semantic contexts and help learners understand when and how to use them appropriately (the degree of formality) (Panthong, et al., 2023). This kind of study is not only academically valuable but also practical, providing insights that are likely to influence language teaching strategies and materials (Li & Fu, 2015; Liu, 2010; Yu, 2023). Moreover, the results from a corpus-based analysis can offer concrete evidence of how language is used in various contexts (Lertcharoenwanich & Phoocharoensil, 2022). Such evidence is invaluable for developing language materials that are both authentic and effective (Phoocharoensil, 2020a, 2020b, 2021a).

Innumerable research has explored the different sets of near synonyms. Early works investigate the general word lists such as 'type, kind and sort' (Panthong, et al., 2023), 'normal, regular, common, and ordinary' (Sangseekaew, 2023), 'provide vs offer'

(Hongxia, 2020), ‘increase vs rise’ (Phoocharoensil & Kanokpermpoon, 2021; Uba & Irudayasamy, 2023), ‘cease, halt and stop’(Chaengchenkit, 2023), ‘blank, empty and vacant’ (Lertcharoenwanich, 2023), ‘error, fault and mistake’ (Phoocharoensil, 2020b), and ‘join, participate and attend’ (Niwesworakarn, et al., 2023). Despite this comprehensive body of literature, the trend of research studies has now been shifted to near synonyms in academic vocabulary. For example, Lertcharoenwanich and Phoocharoensil (2022) explored the synonyms of Purpose, Goal and Objective which were included in the Academic Word List (AWL) by Coxhead (2000). Later, the vocabulary researchers follow this tradition by researchers some sets of words such as ‘assess, evaluate, and measure’ (Sridhanyarat & Phoocharoensil, 2023), and ‘restrict vs constrain’ (Narkprom, 2024).

To date, research focusing on near synonyms in academic wordlists remains relatively scarce. Academic Word List (AWL) by Coxhead (2000), one of the most renowned wordlists, includes the words 'consist of' and 'comprise,' but notably omits 'compose.' As presented in Table 1 below, these three words are defined similarly in major online dictionaries such as the Oxford Learner’s Dictionaries, Cambridge, and Longman dictionaries, where they convey the concept of including various elements such as things or people. However, while Cambridge online resources detail the grammatical usage of these words (Consist, Comprise, or Compose?, 2024), descriptions of their contextual application are limited. For this reason, some language learners might encounter some difficulties using these three words. The selection of these words for focused study is based on two key reasons. First, as these target words in the AWL list, Lertcharoenwanich and Phoocharoensil (2022) posited that researching near synonyms of academic words could possibly provide substantial benefits for the classroom since the results of corpus-based study emerge a piece of solid evidence of real language use. Second, vocabulary specialists (Baumann & Graves, 2010; Nagy, et al., 2012; Warnby, 2023) suggested academic word list serves as the key element to engage learners to communicate effectively and successfully in an academic circle.

Table 1 The Meanings of three target words in three online dictionaries

Words	Oxford Dictionary	Cambridge	Longman
consist	to be formed from the things, people, or activities mentioned	to be made of or formed from something	to be formed from two or more things or people

Table 1 (continue)

Words	Oxford Dictionary	Cambridge	Longman
comprise	to have somebody/ something as parts or members	to have things or people as parts or members; to consist of	to consist of particular parts, groups etc
compose	compose something (formal) to combine together to form a whole	to be the parts that something is made of	be composed of something to be formed from a number of substances, parts, or people

Despite shortcomings, this study articulated two research objectives below. It is hoped that the results of this study will inform the use of these three near synonyms for second language writers who aims to write a piece of academic writing. In addition to that, some language teachers could implement the results of differences and their semantic preferences to the learners while some material maker could get some benefit from the results to compile evidence-based material for learners. The following parts introduces the objective, methodology and results, respectively.

Objectives

1. To explore the differences in term of the degree of formality among three academic near synonyms, namely consist, comprise and compose
2. To investigate the noun, and adverb collocations in relation to semantic preferences of consist, comprise and compose

Methodology

Target Words

In line with the research objectives, this study selected three academic words—'consist,' 'comprise,' and 'compose'—from the COCA website. These words were chosen because they are part of an academic word list. Researchers (Lertcharoenwanich & Phoocharoensil, 2022; Narkprom, 2024) suggested that analyzing near synonyms in the Academic Word List (AWL) could significantly benefit vocabulary instruction.

Data Collection

The data collection of this present study was considered as the corpus-based analysis approach. The following paragraphs explain why this study relied on available online corpus and how the data retrieved from COCA was collected.

The data collection of target words in this present study was obtained from Corpus of Contemporary American English (COCA). This corpus is the collection of a wide range of language use such as academic, news, spoken, website, blog, etc. Apart from that, COCA is considered as the big, available corpus because it includes 1.1 billion words used in the United States from 1990 to 2019 (Davies, 2019). Thus, this variation of language could be aligned with the framework of Phoocharoensil's (2020a) indicating the degree of formality of target words. In addition to that, the use of COCA could potentially inform and raise the student's awareness of writing a piece of academic text. For example, Liontas, et al., (2023) experimented with the students to use the COCA corpus in the academic writing, finding that the students have improved their academic writing and skills of collocation in a particular genre. Lastly, this corpus is freely available as the ready-to-use corpus, so this corpus allows the researcher to access from the real language use.

The data collection involved the following steps. In the first step, each target word was searched on the website. After that, the research arranged the frequency of each target word in each genre, so this would allow the researchers to answer the first research question. Then all of the results presented in Browse were transferred to an Excel file. In the second step, each target word was typed and searched in the Word menu bar. This function allows the researchers to access the collocation menu presenting the frequency and MI scores. Phoocharoensil (2021) raised the concern that the frequency of collocation may not sufficiently reflect the semantic preferences, so MI scores are another factor in determining whether the words are collocated (Chung, 2011). Even though the cut-off point of MI score is very arbitrary, Chung (2011) generalized that MI scores should be set at ≥ 3 . Moreover, according to the research objective, this present study explores three verbs. Only noun and adverb collocations are collected and analyzed since it is the nature of verbs and their collocation based on the notion of vocabulary researchers (Chaengchenkit, 2023; Narkprom, 2024; Sridhanyarat & Phoocharoensil, 2023).

Data Analysis

An analysis of the data relies on Phoocharoensil's (2021) framework. This framework introduces to two-step method to distinguish the differences of near synonyms: the degree of formality and the analysis of their semantics preferences. The first analysis process is to investigate their degrees of formality. In relation to the Excel file in the data collection, the arranged list of frequency in all genres are used to explain their nature of occurrence in language. The second analysis process is to analyze their semantics preferences. Despite the list of collocations, Phoocharoensil (2020a, 2021a, 2022) proposes that limiting the analysis to only 30 collocations would enable researchers to explore the results more clearly and streamline the analytical process. Therefore, only the top 20 collocations of each word were analyzed using the thematic analysis. After that, the analysis is presented in the categories of words.

Results

Together with the data collection and data analysis, this present study introduces two major research results. The following subsections are dedicated to reporting the research results aligned with the research objectives.

Differences among target words

According to the first research objective, the results found that all three near synonymies are prevalent in the academic genre and in written genres including website, magazine, and blog (see Table 2). Notably, 'compose' appears frequently in the academic category within the COCA corpus, registering the lowest usage across all examined genres despite its classification in the AWL. However, these three target words are relatively rare in spoken genres (spoken and TV/M) and fiction. For example, the word 'consist' appears in the spoken genre 111 times and TV/M 78 times, and the word 'comprise' occurs in fiction 57 times and TV/M for 34 times. Similar to 'consist', the word 'compose' is relatively rare because it appears in TV/M for 132 times and in spoken discourse for 121 times.

Table 2 The distribution of frequency and genre of three target words

Consist			comprise			compose		
genre	frequency	per million	genre	frequency	per million	genre	frequency	per million
ACAD	1737	14.5	ACAD	1278	10.67	ACAD	915	7.64
WEB	1300	10.46	WEB	575	4.63	MAG	463	3.67

Table 2 (continued)

Consist			comprise			compose		
genre	frequency	per million	genre	frequency	per million	genre	frequency	per million
MAG	753	5.97	MAG	529	4.2	FIC	461	3.9
BLOG	686	5.33	BLOG	412	3.2	WEB	444	3.57
NEWS	385	3.16	NEWS	321	2.64	BLOG	306	2.38
FIC	146	1.23	SPOK	77	0.61	NEWS	188	1.54
SPOK	111	0.88	FIC	57	0.48	TV/M	132	1.03
TV/M	78	0.61	TV/M	34	0.27	SPOK	121	0.96
Total	5196		Total	3238		Total	3030	

Semantic preferences

According to the second research objective, the result manifested that these target words show their semantic preferences. The following subsections show the quantitative results (frequency and MI scores) and qualitative results (categories and concordance lines).

Noun collocates and Semantic Preferences

The first type of collocates that reflects the semantic references is a noun. Table 3 below reports data on the frequency and MI values of specific nouns used with the verbs "consist," "comprise," and "compose." For the verb "consist," nouns such as "item," "example," and "basis" are frequently used, with "item" occurring in the highest frequency at 713 times and a notable MI score of 4.82. In the category of "comprise," "element," "group," and "part" are among the top nouns. "Element" stands out with the highest frequency, highlighting its importance in contexts where "comprise" is used to denote the parts of a whole, indicating a comprehensive inclusion. For "compose," "music," "song," and "piece" lead in frequency, with "music" having a significantly high frequency and M value. This aligns with the common use of "compose" in musical contexts, where it is typically associated with the creation of musical pieces. Across all three verbs, there is a clear pattern of noun usage that aligns with the semantic properties of the verbs themselves. "Consist" tends to collocate with nouns that signify components, "comprise" with nouns that suggest inclusion, and "compose" with nouns related to creation, particularly in artistic contexts.

Table 3 Noun collocates of consist, compose, and compose

NOUN									
Rank	Consist			Comprise			Compose		
	Noun collocate	Frequency	MI value	Noun collocate	Frequency	MI value	Noun collocate	Frequency	MI value
1	item	713	4.82	percent	636	3.97	music	681	4.41
2	sample	686	5.12	group	575	3.42	song	332	4.15
3	series	307	3.01	item	258	4.61	letter	208	3.44
4	set	277	3.25	sample	251	4.93	piece	207	3.12
5	scale	267	3.85	population	176	3.51	poem	166	5.31
6	participant	231	3.38	individual	151	3.4	committee	149	3.3
7	component	220	4.11	majority	143	3.79	text	136	3.43
8	unit	192	3.06	committee	120	3.33	item	132	3.31
9	diet	181	4.11	set	112	3.2	element	118	3.37
10	survey	172	3.21	element	98	3.44	score	105	3.14
11	questionnaire	172	5.46	category	98	3.88	sample	88	3.08
12	session	155	3.29	unit	92	3.26	e-mail	82	3.79
13	phase	154	3.93	component	79	3.89	melody	78	5.98
14	layer	154	4.02	species	73	3.14	representative	74	3.65
15	panel	151	3.46	scale	73	3.24	works	72	3.33
16	pair	123	3.03	panel	72	3.65	layer	72	3.84
17	instrument	118	3.39	representative	67	3.85	panel	68	3.22
18	sequence	104	3.71	minority	66	3.59	symphony	66	5.7

Table 3 (continued)

NOUN									
Rank	Consist			Comprise			Compose		
	Noun collocates	Frequency	MI value	Noun collocates	Frequency	MI value	Noun collocates	Frequency	MI value
19	representative	97	3.12	portion	63	3.94	poetry	61	4.22
20	subscale	97	6.05	variable	55	3.41	particle	59	4.32

Together with the selected 30 noun collocation of each word to analyze them into themes, this present study found that there are six themes for all words including Group. Component, Measurement, Music, Literary work and Research instrument. The following paragraphs explain the noun collocations based on their semantic preferences.

Table 4 Themes of examples noun collocates of consist, compose and compose

Category	Consist	Comprise	Compose
Group	Cluster, set, series, string, mixture, cluster, pair	Group, species, sample, minority, majority, portion, bulk	
Structural elements	Component, unit, module, molecule, layer	Item, unit, component, layer, panel, atom, variable	Atom, carbon, particle, hydrogen, layer
Measurement	scale, subscale	scale, subscale, acre	
Music			music, song, melody, symphony, opera, verse, musician
Literary Works			poem, text, e-mail, poetry, sentence, essay, narrative

Table 4 (continued)

Category	Consist	Comprise	Compose
Research Instrument	questionnaire, instrument, survey	questionnaire	instrument

The word "consist" is frequently found with nouns that belong to categories such as group, structural elements, measurement, and research instrument. This indicates a common use of "consist" in the context where the components of a whole are listed or explained in detail. Notably, the most populous category under "consist" is a group such as set, component, and series (see Example 1). Moreover, the noun collocations can be also followed by structural elements such as component, unit, module, and the scientific term like molecules (Example 2). The third theme of noun collocated with verb consist is the measurement which consists of scale and subscale (Example 3). Finally, the research instrument is considered as the last theme of this noun collocation. It consists of the words explaining the research tool to collect the data: questionnaire, instrument and survey (Example 4).

1. Our series consisted of 12 cases which were classified according to anatomic criteria ...
2. The brain's self-made cannabis consists of molecules called endocannabinoids, which hit the targets.....
3. The scale consists of 15 semantic differential items encompassing the dimensions of (1) attractiveness (attractive/unattractive)
4. ... and try to answer why, and not just how much. The survey consisted of 19 open and closed type questions (circling answers

The word "comprise" has a similar themes to "consist" in its function of outlining the elements of a whole, yet they share somewhat different collocations such as minority, majority, and portion. First, the collocations signifying the group include group, sample, minority, and species (Example 5). Second, the structural elements consist of the collocations, namely component, layer, panel, and atom (Example 6). Third, the theme of measurement is similar to the word 'consist', and the word also consists of scale, subscale, and acre (Example 7). The last one refers to the research instrument including the questionnaire (Example 8).

5. *Aurantimonas coralicida*, an obligate aerobe, belongs to the order Rhizobiales. The family comprises only two species, they are recognized because they have a bacillus....

6. A potentially supporting piece of information was recognized around 1980. The interstellar material comprises mostly atoms (with some molecules).

7.it is as interesting a place as one would find in that country. It comprises about four hundred acres of land, and is inhabited by one man

8. Our questionnaire comprises of 12 such statements among which 10 is positive and 2 are negative statements.

In contrast to “consist” and “comprise”, “compose” is distinctly associated with creative and construction-related settings, as evidenced by its connection to categories like music and literary pieces. Nouns such as music, song, and poem emphasize “compose’s” involvement in artistic endeavors (Example 9). In the same token, some noun such as poem, text, and poetry are classified in a literary work (Example 10). “Compose” also shares the category of structural elements (Example 11) and research instrument (Example 12) with the other two verbs, indicating its widespread application in detailing both functional and scientific assemblies.

9. Twelve percent reported having rather extensive composing experience, such as writing many songs, composing a string quartet, and composing a solo piece for...

10. topic before when we were engaged in writing fiction and expository texts and also when composing poetry to a lesser extent, but we wondered how memoirists ...

11. Polyfluorinated iodine alkanes (PFIs) are organic iodides composed of a fluorinated carbon backbone terminated by iodine substitution (Table 1)...

12. First, when composing a quality appraisal instrument, researchers need to decide what the focus of that quality appraisal should be.

In summary, each verb has its specific communicative roles. The word “consist” outlines the components of collections or systems while “comprise” highlights the proportions within a whole, and “compose” concentrates on the genesis of both artistic and structural formations.

Adverb collocates and Their Semantic Preferences

The second type of semantic preference refers to the adverb collocation. Table 5 provided the frequency and MI scores of adverbs collocated with the verbs “consist,” “comprise,” and “compose.” It highlights the distinct preferences for adverb usage with these verbs, and it illustrates the ways in which they are applied in real contexts. For example, the adverb “invariably” is most frequently used with “consist,” reflecting a strong collocation, while

"mostly" is predominantly linked with "comprise," and "largely" appears significant for "compose." These patterns suggest that certain adverbs are more apt in specific contexts, likely due to the subtle differences in the meanings or typical usage scenarios of these verbs.

Table 5 Adverbs collocates of consist, compose and compose

ADVERB									
Rank	Consist			Comprise			Compose		
	adverb collocater	Frequency	MI value	adverb collocater	Frequency	MI value	adverb collocater	Frequency	MI value
1	mostly	487	4.63	mostly	105	3.67	entirely	237	5.06
2	mainly	411	5.92	approximately	85	4.54	mostly	189	4.18
3	primarily	357	5.31	mainly	81	4.84	primarily	166	5.12
4	entirely	311	4.53	entirely	73	3.7	largely	149	4.49
5	largely	286	4.51	largely	73	3.8	mainly	126	5.14
6	solely	149	5.27	primarily	65	4.11	carefully	81	3.71
7	typically	127	3.49	wherein	38	6.37	exclusively	53	4.83
8	essentially	85	3.08	roughly	32	3.32	originally	38	3.18
9	approximately	82	3.23	solely	26	4	newly	36	3.41
10	exclusively	71	4.33	exclusively	23	3.96	solely	36	4.14
11	chiefly	48	6.17	predominantly	22	4.83	predominantly	36	5.21
12	principally	46	6.11	overwhelmingly	13	4.16	beautifully	33	4.77
13	predominantly	29	3.98	e.g.	8	4.65	overwhelmingly	15	4.02
14	wholly	20	3.46	collectively	7	3.22	principally	15	5.42
15	i.e.	15	3.13	i.e.	6	3.07	wholly	12	3.64
16	wherein	15	3.77	chiefly	6	4.42	meticulously	12	5.34

Table 5 (continued)

ADVERB									
Rank	Consist			Comprise			Compose		
	adverb collocate	Frequency	MI value	adverb collocate	Frequency	MI value	adverb collocate	Frequency	MI value
17	e.g.	9	3.56	disproportionately	5	4.1	collaboratively	11	6.23
18	predominately	6	4.72	principally	4	3.85	i.e.	10	3.46
19	vertically	5	3.02	predominately	4	5.4	specially	10	3.85
20	ethnically	5	3.34	geographically	3	3.45	chiefly	10	4.82

The analysis of adverb collocation preferences of the verbs "consist," "compose," and "comprise" reveals different categories that emphasize various aspects of usage and meaning. These categories include the degree of composition, approximation, specific conditions or exceptions, descriptive or qualitative attributes, and collective inclusion. Each verb shows a preference for certain adverbs that enhance its specific semantic context. The explanation of each target word are as follows.

Table 6 Themes of adverb collocates of consist, compose and compose

category	Consist	Comprise	Compose
Degree of Composition	mostly, mainly, primarily, entirely, largely, solely, exclusively, chiefly, principally, predominantly, wholly, predominately	mostly, mainly, entirely, largely, primarily, solely, exclusively, predominantly, chiefly, principally, predominately, overwhelmingly, disproportionately	entirely, mostly, primarily, largely, mainly, solely, exclusively, predominantly, overwhelmingly, chiefly, principally, wholly

Table 6 (continued)

category	Consist	Comprise	Compose
Approximation	approximately	approximately, roughly	
Specific Conditions or Exceptions	i.e., wherein, e.g.	wherein, e.g., i.e., therefore	i.e., e.g., thereon
Description	ethnically, electrically, sequentially, vertically, steeply, irregularly	ethnically, geographically, functionally, densely, ethically, chemically, regionally, colorfully	carefully, meticulously, beautifully, artfully, thoughtfully, hastily, loosely, originally, newly, specially, digitally, radially, visibly, collaboratively, outwardly
Collection		collectively	

For the word "consist," the adverbs are used primarily under the degree of composition with terms like "mostly" and "entirely," signifying the extent of inclusivity within a subject (Example 13). An approximation is noted distinctly with "approximately" (Example 14) while specific conditions or exceptions are marked by abbreviations like "i.e." (Example 15). The descriptive or qualitative category is populated with adverbs that describe the manner or condition of the composition, such as "sequentially" and "vertically" (Example 16).

13. engines are essentially massive databases that cover wide swaths of the Internet. They mostly consist of the parts of at least one program called spider, or crawler or bot...

14. ... at the N-terminus, that are lacking in the Class I enzymes which are consisted of approximately 390 amino acids. There is also another short (10-14 residue)

15. - the fact that two of the failed anterior FDPs were 'experimental' (i.e. consisted of 5 units) and, therefore, probably more liable to fracture. In....

16. Etienne-Nugue and Laget 1985:172-83). The designs, mostly on native cloth, consist of repetitive vertically or horizontally organized, black abstract motifs, and occasional painted figures

In the case of "comprise," the categories are similarly structured. The degree of composition also includes "mostly" and "disproportionately," pointing to a broader range of

intensity (Example 17). Approximation extends to "roughly," (Example 18) and collective inclusion is introduced with "collectively," emphasizing a group or total sum perspective (Example 19). The qualitative adverbs expand to include diverse aspects such as "densely" (Example 20) and "chemically," broadening the scope of the description.

17. to remind us that war is anything but escapable. Preceded by a chapter mostly comprised of an internal monologue by a sexually repressed

18. NATO forces or International Security Assistance Forces (ISAF).11 Given that the Pashtuns comprise roughly 30 to 40 percent of Afghanistan's population....

19. unfamiliar to most illiterate people " (Hirsch 136). All the terms collectively comprise an operational definition of cultural literacy. Hirsch claims that being culturally literate is "

20.analysis for this study are the Census Urban Areas (CUAS), which comprise contiguous, densely populated regions with a minimum of 50,000 residents and a density of

The themes of the verb 'compose' reveal the degree of composition as the major themes such as entirely, mostly, primarily (Example 21), followed by a variety of descriptive adverbs like "originally" (Example 22). Specific conditions or exceptions which include i.e., e.g., and thereon are considered as the smallest proportion of all themes (Example 23). This verb's semantic preferences suggest a focus on the quality and method of inclusion within a composition.

21. 2003, less than two months after the Hutton Inquiry finished hearing testimonies, is composed entirely of direct quotes from the transcripts of the Hutton Inquiry,

22. the diffuser 180 having a plurality of dot elements 182 which are formed thereon and compose a radially variable dot specific pattern 196, the radially variable

23. Explain that the music used as a sound track for Time Flies was not originally composed for the book. Share information about sound tracks and a music editor's role....

In conclusion, the semantic preferences of "consist," "compose," and "comprise" highlight specific linguistic patterns that reflect their unique contextual applications. All of three words may emerge the themes including the degree of composition, specific Conditions or exceptions, and the description, but as seen in Table 6, they partially share the similar adverb collocations. In addition to that, only the words 'consist' and 'compose' are used to express the approximation of the components while only the word 'comprise' can be used to show the sense collection. To this end, these words cannot be used interchangeably.

Discussion

The main purposes of this study are to explore the differences among three target words and investigate their semantic preferences. There are two major findings regarding the analysis as discussed in the following paragraphs.

In the first place, the result found that three target words are common in the written register especially academic discourse while they are less common in the spoken counterpart. In addition, the word 'compose' is prevalent in the academic genre despite the inclusion of AWL. For the nature of genres, this present study shares the similarities with previous studies (Kruawong & Phoocharoensil, 2022; Lertcharoenwanich & Phoocharoensil, 2022; Narkprom, 2024; Sridhanyarat & Phoocharoensil, 2023). The prevalence of the words in different written genres could be explained by the fact that the words included in AWL are more common in written language than verbal ones. For the word 'compose', it could be assumed that AWL was first created in 2000, and to date, it has been almost three decades. It might be related to some language evolution.

In the second place, an analysis of the semantic preferences found that these three words have their own specific preferences despite the sharing collocations, so it can be concluded that these three cannot be used interchangeably. The results of this study are in line with the previous research (Hongxia, 2020; Lertcharoenwanich, 2023; Lertcharoenwanich & Phoocharoensil, 2022; Lin & Chung, 2021; Liu & Espino, 2012; Niwesworakarn, et al., 2023; Xiong & Liu, 2022). In spite of same themes of semantic preferences of collocation, it should be taken into the consideration that all the words should be meticulously selected to use in each context appropriately. This could inform the pedagogical implication that the language teachers should present the differences among these three words to students to improve their writing and speaking skills.

Recommendation

The findings of the current study have several practical implications for pedagogical contexts. Firstly, the insights derived from a corpus-based analysis of genre-specific variations among three target words can be integrated into instructional practices within the classroom. This approach not only facilitates genre awareness but also enhances students' understanding of semantic preferences, enabling them to select words that are contextually appropriate for their academic writing. Secondly, researchers and graduate students preparing to publish scholarly articles or compose their theses may find these findings particularly beneficial. The

corpus-based evidence presented surpasses traditional dictionary definitions, offering more nuanced understandings of word usage (Phoocharoensil, 2021b). Lastly, the study's results offer valuable resources for creators of educational materials (Lertcharoenwanich, 2023; Panthong, et al., 2023; Sridhanyarat & Phoocharoensil, 2023). These practitioners can implement the findings as examples or suggestions to illustrate the practical application of the three target words. Furthermore, the results can be used to develop exercises aimed at bolstering learners' writing skills.

This present study recommended the following aspects for the future research. First, this present study obtained the data from COCA corpus, so the variety of Englishes is rare. The future research should use the different available corpus to explore the variety of Englishes as well as the near synonyms. Secondly, this study did not assign the position of collocation. For example, Kruawong & Phoocharoensil's (2022) work explored only post-occurring words of differences among near synonyms, presenting the way in which writers can use in their pieces of writing. Lastly, this present study adopted only two criteria including the degree of formality and semantic preferences to analyze the near synonyms. Given this limitation, the future work might adopt other criteria to distinguish these differences.

References

- Baumann, J.F. & Graves, M.F. (2010). What Is Academic Vocabulary?. **Journal of Adolescent & Adult Literacy**, 54(1), 4-12.
- Cambridge Dictionary. (2024). **English Dictionary, Translations & Thesaurus**. [Online]. Available : <https://dictionary.cambridge.org/> [2024, March 23].
- Chaengchenkit, R. (2023). A Corpus-based study of the synonyms cease, halt, and stop. **LEARN Journal : Language Education and Acquisition Research Network**, 16(1), 473-494.
- Chung, S.F. (2011). A Corpus-based Analysis of “Create” and “Produce.” **Chang Gung Journal of Humanities and Social Sciences**, 4(2), 399-425.
- Consist, comprise or compose. (2024). In **Cambridge Grammar**. [Online]. Available : <https://dictionary.cambridge.org/grammar/british-grammar/consist-comprise-or-compose> [2024, March 23].
- Coxhead, A. (2000). A New Academic Word List. **TESOL Quarterly**, 34(2), 213.
- Davies, M. (2019). **The Corpus of Contemporary American English (COCA): One billion words, 1990-2019**.

- Hongxia, M. (2020). A Comparative Analysis of Verb Synonyms from the Perspective of Valency Structure in the Context of "Provide and Offer." **IRA International Journal of Education and Multidisciplinary Studies**, 16(4), 217-223.
- Sangseekaew, K. (2023). A Corpus-Based Study of English Synonyms : Normal, Regular, Common, Ordinary, and Usual. **The New English Teacher**, 17(1), 113-135. [In Thai]
- Kruawong, T. & Phoocharoensil, S. (2022). A Genre and Collocational Analysis of The Near-Synonyms Teach, Educate and Instruct: A Corpus-Based Approach. **Teflin Journal**, 33(1), 75-97.
- Lertcharoenwanich, P. (2023). Analysis of Collocations and Semantic Preference of the Near-synonyms: Blank, Empty, and Vacant. **LEARN Journal : Language Education and Acquisition Research Network**, 16(1), 365-383. [In Thai]
- Lertcharoenwanich, P. & Phoocharoensil, S. (2022). A Corpus-based Study of the Near-synonyms: Purpose, Goal and Objective. **REFlections**, 29(1), 148-168. [In Thai]
- Li, Y.F. & Fu, J.M. (2015). **A Study on the Corpus-Based Teaching and Learning of English Near Synonyms**. Proceedings of 2015 International Symposium-College Foreign Languages Education Reform and Innovation.
- Lin, Y.Y. & Chung, S.F. (2021). A Corpus-Based Study on Two Near-Synonymous Verbs in Academic Journals : Propose and Suggest. **English Teaching and Learning**, 45(2), 189-216.
- Liontas, J.I., Bangun, I.V. & Li, S. (2023). Lexical Collocational Instruction in EAP Writing Via COCA. **Teaching English with Technology**, 23(1), 80-106.
- Liu, D. (2010). Is It A Chief, Main, Major, Primary, or Principal Concern? : A Corpus-Based Behavioral Profile Study of the Near-Synonyms. **International Journal of Corpus Linguistics**, 15(1), 198-228.
- Liu, D. & Espino, M. (2012). Actually, Genuinely, Really, and Truly: A Corpus-Based Behavioral Profile Study of Near-Synonymous Adverbs. **International Journal of Corpus Linguistics**, 17(2), 198-228.
- Longman Dictionary of Contemporary English(LDOCE). (n.d.). **Dictionary**. [Online]. Available : <https://www.ldoceonline.com/> [2024, March 25].
- Nagy, W., Townsend, D. & Schmitt, N. (2012). Words as tools: Learning academic vocabulary as language acquisition. **Reading research quarterly**, 47(1), 91-108.

- Narkprom, N. (2024). A Corpus-Based Investigation into English Synonymous Verbs ‘Restrict’ and ‘Constrain.’ **LEARN Journal : Language Education and Acquisition Research Network**, 17(1), 688-714. [In Thai]
- Niwesworakarn, N., Phoocharoensil, S., Petkaew, J. & Intasingh, S. (2023). Collocational Patterns of the Synonyms Join, Participate, and Attend: A Corpus-Based Study. **The International Journal of Communication and Linguistic Studies**, 22(1). 63-77.
- Oxford Learner’s Dictionaries. (n.d.). **Find definitions, translations, and grammar explanations at Oxford Learner’s Dictionaries.** [Online]. Available : <https://www.oxfordlearnersdictionaries.com/> [2024, March 23].
- Panthong, P., Chutopama, N. & Putklang, T. (2023). Type, Kind, or Sort: Exploring Near Synonyms through Corpus Analysis. **Academic Journal of Buriram Rajabhat University**, 15(2), 50-69. [In Thai]
- Petcharat, N. & Phoocharoensil, S. (2020). A Corpus-Based Study of English Synonyms: Chance and Opportunity. **REFlections**, 27(2), 218-245. [In Thai]
- Phitayakorn, W. (2016). **A Corpus-Based Study of English Near Synonyms : Advise, Recommend, and Suggest.** Independent Study Thammasat University. [In Thai]
- Phoocharoensil, S. (2020a). A Genre and Collocational Analysis of Consequence, Result, and Outcome. **3L The Southeast Asian Journal of English Language Studies**, 26(3), 1-16.
- Phoocharoensil, S. (2020b). Collocational patterns of the near-synonyms error, fault, and mistake. **International Journal of Communication and Linguistic Studies**, 19(1), 1-17.
- Phoocharoensil, S. (2021a). Multiword units and synonymy: Interface between collocations, colligations, and semantic prosody. **GEMA Online Journal of Language Studies**, 21(2), 28-45.
- Phoocharoensil, S. (2021b). Semantic prosody and collocation: A corpus study of the near-synonyms persist and persevere. **Eurasian Journal of Applied Linguistics**, 7(1), 240-258.
- Phoocharoensil, S. (2022). Primary, Main, and Major: Learning the Synonyms through Corpus Data. **GEMA Online Journal of Language Studies**, 22(4), 76-89.
- Phoocharoensil, S. & Kanokpermpoon, M. (2021). Distinguishing the near-synonyms ‘increase’ and ‘rise’: Genre and collocation investigation. **Kasetsart Journal of Social Sciences**, 42(4), 968-975. [In Thai]

- Sridhanyarat, K. & Phoocharoensil, S. (2023). A Corpus-based Investigation of English Near-synonyms : Assess, Evaluate and Measure. **Humanities, Arts and Social Sciences Studies**, 23(1), 208-219. [In Thai]
- Taylor, J.R. (2003). Near synonyms as co-extensive categories: “High” and “tall” revisited. **Language Sciences**, 25(3), 263-284.
- Uba, S.Y. & Irudayasamy, J. (2023). Is it ‘Increase’ or ‘Rise?’ A Corpus-based Behavioural Profile Study of English Near-Synonym Verbs. **Mextesol Journal**, 47(1), 1-8.
- Warnby, M. (2023). Academic vocabulary knowledge among adolescents in university preparatory programmes. **Journal of English for Academic Purposes**, 61, 101203.
- Xiong, Y.H. & Liu, D.F. (2022). A Corpus-based Analysis of English Near-synonymous Adverbs: Absolutely, Utterly. **Journal of Literature and Art Studies**, 12(4), 359-365.
- Yu, L. (2023). Exploring Semantic Nuances in English Synonyms through Multimodal Analysis. **English Language Teaching and Linguistics Studies**, 5(4), 71-81.